Zeitschrift für Arbeitswissenschaft (2024) 78:349–358
https://doi.org/10.1007/s41449-024-00436-3

WISSENSCHAFTLICHE BEITRÄGE

Human-centered approaches in working environments needs ethical reflection—Experiences from KARL

Bettina-Johanna Krings[1] · **Philipp Frey**[1]

Accepted: 5 August 2024 / Published online: 3 September 2024

Abstract

In our paper, we discuss some insights from our work in the regional Centre of Competence of Artificial Intelligence (KARL), which deals with the implementation and application of Artificial Intelligence in working and learning environments. We present a conceptual framework that we have developed to define social, legal and ethical dimensions of AI in working environments. As a first step, the article describes the definition of these aspects within the technical development processes in KARL with an emphasis on ethical issues. Furthermore, a practical example is used to illustrate how normative premises can be elaborated to inform the design process of AI-systems. Next, the process of development and implementation of AI-systems in working environments is taken into account. The article explicitly emphasises the importance of ethical reflection, i.e. on norms such as fairness, social sustainability and the creation of meaningful work, to inform in human-centered approaches to AI-based work.

Practical Relevance: This paper discusses a process-oriented, human-centred approach to AI-based work. The joint negotiation and definition of "human-centred work with AI" in each specific context with all concerned stakeholders lies at its heart. The concretisation of these implications is of great importance for the practical implementation of the overarching concept, as it is only in practice that it becomes clear whether and how a participatory design approach is successful, which areas of tension and problems arise and how these can be dealt with.

Keywords Artificial Intelligence · Socio-technical working environments · Participation · Ethics

Die Gestaltung von menschzentrierten Ansätzen in sozio-technischen Arbeitsumgebungen benötigt ethische Reflexion – Erfahrungen von KARL

Zusammenfassung

Im vorliegenden Artikel werden erste Ergebnisse aus dem Projekt KARL vorgestellt, die sich auf die Gestaltung und Implementierung von KI-Systemen in Arbeitsumgebungen beziehen. Für diesen Zweck wurde ein Rahmenkonzept entwickelt, das ethische, rechtliche und soziale Aspekte („ELSA") definiert und in einen breiten Entwicklungsprozess im Rahmen des Projektes einbindet. Der Artikel beschreibt in einem ersten Schritt die Konkretisierung dieser Aspekte im Rahmen der Gestaltungsprozesse in KARL, wobei der Schwerpunkt auf den ethischen Implikationen liegt. In einem zweiten Schritt werden auf der Basis eines Anwendungsbeispiels Prozesse in den Blick genommen, die grundlegende Fragen zu Themen wie etwa Fairness und Gestaltung der Arbeitsinhalte berühren. Vor dem Hintergrund ethischer Betrachtungen, werden diese Fragen zunehmend mehr auf der Ebene sozialer Normen im Hinblick auf die Ausgestaltung zukünftiger Arbeitswelten diskutiert.

✉ Dr. Bettina-Johanna Krings
Bettina-Johanna.Krings@kit.edu

[1] Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Praktische Relevanz: Der Artikel stellt einen prozessorientierten Ansatz zur Entwicklung und Einführung von KI-Systemen in Arbeitsumgebungen zur Diskussion. In dessen Zentrum steht die gemeinsame Aushandlung und Definition „menschzentrierter Arbeit mit KI“ im jeweiligen konkreten Kontext mit allen betroffenen Akteuren. Die Konkretisierung dieser Implikationen ist für die praktische Furchtbarmachung des übergeordneten Konzepts von großer Bedeutsamkeit, denn nur in der praktischen Anwendung zeigt sich, ob und wie partizipativ angelegte Gestaltung gelingt, welche Spannungs- und Problemfelder auftreten und wie diese bearbeitet werden können.

Schlüsselwörter Künstliche Intelligenz · Gestaltung sozio-technischer Arbeitswelten · Partizipation · Ethik

1 Artificial Intelligence in Working Environments—an Entroduction

For decades, digital transformation has been taking place in every societal field. Production and working processes rely on data-based systems. Based on technological innovations “digitalizing production is gaining a new level of quality [...]: The Internet of Things, machine-to machine communication and manufacturing facilities that are becoming ever more intelligent are heralding a new era” (BMWK 2024). Additionally, the organization and coordination of daily life can no longer be imagined without the use and the application of data and digital end devices. However, in recent years, controversial public debates on Autonomous Driving or the use of ChatGPT in education (Albrecht 2023) have almost overshadowed the ongoing processes of digitization across all industries and sectors. Although Artificial Intelligence (AI)[1] systems have already been successfully introduced in various areas of application, a fundamental critical review of the use of AI has begun in public and scientific debates with political implications (Spiekermann 2019; Seyfarth and Roberge 2017). In industry, AI application is embedded into the technical path of the so-called Industry 4.0, exploring further potential of autonomous systems in production (El-Haouzi et al. 2021; Hirsch-Kreinsen and Karacic 2019). Here, so-called “learning systems” (Hirsch-Kreinsen and Karacic 2019) are developing complex processing data chains in order to create digital representations of production processes and to coordinate or to automatize ever larger parts of them.

For the workers, these technological changes cannot be longer be tackled by “learning by doing” models. In contrast, new human-machine-interactions (HMI) based on AI are provoking new learning and adaptation models, taking into account new forms of HMI as well as the increasing complexity of working environments (Schröter 2019; Krings et al. 2021). At the same time, the permanent collection of data on working flows can be used not only as a basis for partly or fully automated decision-making processes, but also to surveil human work (Christl 2021; Schaupp 2021). Besides traditional risks like job losses or/and decrease and devaluation of human qualification and skills by new technological systems, there are a set of risks which should be avoided when introducing the concept of Industry 4.0 (Hirsch-Kreinsen et al. 2018). In response, the concept of human-centered work was (re-)established in the strategic design of Industry 4.0 in order to make significant progress towards a new quality of industrial work (BMWK 2024).

Likewise, with the rise of AI it is essential to consider new aspects such as new forms of surveillance and control in the workplace or the protection of individual data privacy have to be considered. “From an ethical point of view, this means that from a macrosocial perspective, the challenge lies not in the means available, but in the shaping of a fair social and economic system ...” (Kirchschlaeger 2021, p. 107). This goal seems remarkable, since expectations on these “technological futures” (Grunwald 2022) strengthens the vision of “fair” work and “fair” organizational pattern of work. Thus, that reflection on the future of work becomes an ethical topic, also in the context of the implementation of AI systems in the workplace.

In this paper, we present the idea of human-centered approaches to AI developed in the context of the regional Centre of Competence KARL[2], which deals with the implementation and application of AI in working and learning environments. This framework operates on two levels: on the level of technical systems to be designed and on the level of a holistic development process, which includes transformation issues in a long-term perspective. We developed a guideline, which provides on the one hand social and ethical orientation for shaping future working environments. On the other hand, we argue for the implementation of an agile cooperative change management (Alpers et al. 2023) in order to reflect on quantitative and qualitative changes of future workplaces and to decide on these

[1] See Definition of AI in the Glossary of KARL: https://zenodo.org/records/10779298, 24.03.2024.

[2] KARL (Künstliche Intelligenz für Arbeit und Lernen in der Region Karlsruhe) is one of five Competence Centers, which has been financed by the Federal Ministry of Education and Research. The Centre involves numerous partners from the Karlsruhe University of Applied Sciences (HKA), Karlsruhe Institute of Technology (KIT) and private enterprises from the region. Its objective is to provide interdisciplinary and transdisciplinary perspectives on AI.: https://kompetenzzentrum-karl.de/ueber-karl/, 02.03.2024.

changes inclusively. In the following, this guideline will be presented with a special emphasis on the question, why and how ethical issues should be addressed in these processes.

We argue that the integration of ethical reflection is particularly relevant in this context. Besides the much-debated legal issues such as the protection of privacy or the integral feature of AI in organizational and/or decision-making processes (Kirchschlaeger 2021), AI-systems in work environments may have a profound impact on the organization of work as well as on the long-term quality of people's working lives. Thus, the introduction of these technological systems is once again provoking a need to clarify what "human-centered" actually ought to mean in this context. Furthermore, it seems to be of great importance to figure out which normative values guide the introduction of AI into these context-sensitive environments. As a result, the ethical evaluation of the use of AI raises awareness of how human actions and experiences will be reshaped by new technologies at work in the nearest future.

2 Ethical considerations on the use of artificial intelligence

> "Ethics has a saying in matters related to technique simply because technology is part of the exercise of human power, namely a form of action, and all human actions or conducts are subjects of moral assessments ..."[3] (Hans Jonas 1985, p. 42)

In Germany, the idea of "human-centered" organization of work dates back to the 1980s, when a huge political program called "*Humanisierung des Arbeitslebens (Humanisation of Work, HdA)*" was established by the former Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung (BMBF)) to respond to the crisis of tayloristic working conditions in industry (Salfer and Furmaniak 1981). In the following decades the program was further developed. In the 1990s new models of work organization such as lean production, group work, job enrichment and increase of qualified work patterns were discussed and established. Here, the program was re-defined in "*Arbeit und Technik*" with a strong focus on technological innovations. In spite of controversial evaluation schemes on the impact of long-term working conditions, there is agreement, that the programs led to important learning processes with regard to human-centered issues at work (Müller 2019). However, in the last decade digitization and KI have raised old and new questions and answers with regard to the organization of work and the future of work more generally (Frey and Osborne 2017).

With regard to the implementation of technological innovations, the creation and modification of working environments in Germany takes place within the framework of the Works Constitution Act (Betriebsverfassungsgesetz) as well as the *Occupational Health and Safety Act* (Arbeitsschutzgesetz). However, as experiences with digitization and the implementation of AI have shown, the interaction of humans and machine evokes new ethical, legal and social challenges (Pereira et al. 2023). There are a number of scientific fields such as Technology Assessment (TA), Ethics of Technology (Grunwald and Hillerbrand 2021) or Ethics-by-Design (Rudschies et al. 2021) that focus on the impact of technological innovations in society, or the co-creation of society and technological development (Grunwald 1994, 2019). In 2010, the discourse on Responsible Research and Innovation (RRI) has raised social, political and ethical issues regarding the philosophical concept of "Responsibility" (Owen et al. 2012) in the field of technological innovation. In the following years, the discourse on RRI guided the European Union's research programs, which included research on the ethical, legal and social aspects ("ELSA") of technological change in order to strengthen potential impacts on environmental protection and societal welfare in European countries.

These discourses have undoubtedly underlined the importance of human responsibility in shaping technological progress, which must also be taken into consideration in the field of AI. Facing these technological systems, "which possess [...] the potential to become similar to humans in certain competences, humans reflect on their own nature" (Kirchschlaeger 2021, p. 42), questions of agency and responsibility seem to become blurred. Profound ethical problems regarding "the attribution and distribution of responsibility [arise] as soon as decisions are delegated to AI systems" (Grunwald 2024, p. 3). Correspondingly, this includes the prominent question of which decisions should not be delegated to technical systems, respectively AI. This question seems to be of high relevance with regard to the premises of the development of future work.

Furthermore, the development of AI is bringing about changes that will profoundly transform the daily lives and work of millions of people. Ethical considerations, here, are manifold, ranging from the definition and clarifications of human autonomy and freedom in different societal context to the evaluation of training programs of technological systems in ethical values and norms, to a macrosocial perspective on the future of work. As a positive example, the German Ethics Council has highlighted as an important criterion of AI that these systems should extend human autonomy and freedom and not limit or even replace them in their contexts (Deutscher Ethikrat 2023). What this means

[3] This quotation was translated into English language by Kirchschlaeger (2021, p. 33).

in concrete terms must, of course, be determined individually for each area of application. Abstract ethical statements do not suffice, as the respective empirical contexts must be adequately considered to acknowledge the ongoing intertwinement of humans and technology in the field of KI (ibid. 2023).

As business models based on AI proliferate, ethical expertise in the field of AI has also developed. This is also the case for working environments, where ELSA criteria have been developed by different actors. For instance, the *High-Level Expert Group on AI* of the European Commission describes its human-centered approach as follows: "[...] To do this, AI systems need to be human-centric resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom." (European Commission 2019, p. 4; Spiekermann 2019; Kirchschlaeger 2021).

These overarching goals are specified in ethical dimensions, which seems considerable with respect to the development of new AI systems in society. According to the experts, this development should "ensure that AI systems are developed, deployed and used in a trustworthy manner" (Ibid., p. 11). The report identifies four dimensions or premises as crucial (Ibid., ff.): (a) Respect for human autonomy; (b) Prevention of harm; (c) Fairness; (d) Explicability. The first issue, respect for human autonomy, in particular is linked to the human-centered design of work environments (Ibid., p. 12), which is described explicitly in the following quotation:

> "Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems. AI systems may also fundamentally change the work sphere. It should support humans in the working environment, and aim for the creation of meaningful work"

Applying this focus on issues such as the preference for AI systems to support and complement human skills and to contribute to meaningful options for human work defines clear premises for AI and HMI design. Furthermore, a human-centered innovation approach should not only be considered in the development stages of the software, but also in the process of its implementation, particularly when it comes to working environments. In this, we follow Welf Schröter in highlighting the qualitative changes that take place when AI is introduced into workspaces (Schröter 2019), both in terms of new challenges introduced for co-determination processes, as well as in terms of the impact that AI can have on the development of capacities and autonomy of workers. Very often, technological changes go unnoticed by the workers because technical innovation is usually initiated, shaped and imposed by employers. In contrast, we will show how a human-centered approach to the implementation of AI can be realized to a large extend through participatory and co-determined processes (Schröter 2019). Although these require time, money and openness on the part of all actors involved, they enable end-users to voice their needs and can thus contribute to a more acceptable and successful implementation of AI.

3 Implementing human-centered approaches to AI in work environments the process-oriented way

3.1 Developing an ELSA-based approach: Methodology

Generally, ELSA-based approaches are embedded in the now well-established and long tradition of interdisciplinary and transdisciplinary research (Mittelstraß 2005; Gethmann 2015). Additionally, it can be understood as a form of "integrated research", a prominent umbrella term (Bellon and Nähr-Wagener 2020) that was fundamentally developed in European research programs such as *Horizon 2020* and the European Program *Responsible Research and Innovation* (RRI). The specific interest of integrated research was to open up scientific disciplines towards societal problems, the transfer of scientific knowledge towards the public and to public participation in research and developing processes of new technologies (Bellon and Nähr-Wagener 2020). In this context, the focus on ethical, legal and social issues was identified in a huge research program (Human genome research in the U.S.) and transferred to European research standards (Bellon and Nähr-Wagener 2020). Furthermore, interdisciplinary and transdisciplinary research has been a constitutive part of Technology Assessment (TA) from the very beginning and the debate on ELSA issues has strongly been integrated into the methodology of TA in recent years (Böschen et al. 2021). Generally, the identification of ELSA-issues is highly context sensitive and needs to take into account the type of technology, its functional demands and its different ethical, legal and social implications with regard to its objectives. Consequently, ELSA issues cannot be easily generalized. Furthermore, the identification of ELSA issues oftentimes leads to a debate on the objectives of technological innovation and (potentially)

to the adaption of innovation processes to the societal value system.

For KARL, the methodology of our ELSA approach included two main steps: (a) the identification of ELSA issues in this specific AI context under the premise of a "human-centered approach"; (b) Generating communication processes with project partners based on the method of stakeholder workshops in order to both involve them in the reflection on AI development and implementation and to integrate these perspectives into the development processes of these technological systems.

Ad a) Identification of ELSA issues in the context of KARL: To facilitate and to guide a process-oriented approach to the human-centered design and implementation of AI, we have developed a set of proposals that are summarized in a White Paper on Ethical, Legal and Social Aspects (ELSA) of AI in the workplace and learning environments (Alpers et al. 2023). The white paper is based on extensive desk research as well as three years of discussions within a (sub-)project team dedicated to ELSA-reflection within KARL. In line with the aim of disciplinary and transdisciplinary cooperation in the project, the team consists of representatives from the following disciplines: Software Engineering, Law, Technology Assessment, Sociology and Philosophy as well as a representative of a trade union network. From the very beginning, the group met regularly to establish a shared structure for communication, work and cooperation. Although ELSA defines different disciplinary angles such as ethical, social and/or legal issues, the group agreed to largely jointly define research topics and expectations, and to jointly resolve the project expectations. Thus, a real ELSA-perspective was developed, which was described in several publications (Alpers et al. 2023, 2024; Krings et al. 2021).

Ad b) In addition to these disciplinary and transdisciplinary discursive processes within our core team, we conducted stakeholder workshops both as part of consortium meetings and as stand-alone events aimed at consortium partners or members of business communities interested in AI. The instrument of the *stakeholder workshop* was chosen to learn about the different societal expectations, the interests, the objectives as well as the individual interests of the participants with regard to the KARL project (Nielsen et al. 2017). The dialogue between the participants should, on the one hand, provide information about the technology and the impact of KI implementation. On the other hand, it seemed important to make visible the relevance of these impacts for the partners involved (Nielsen et al. 2017).

Indeed, these workshops were extremely useful to raise awareness for the relevance of ELSA in the context of the use of AI—and to test out early designs of our instruments. Raising open questions, sharing problem orientation or providing different perspectives with regard to changing HMI in working environments was an important instrument to change the mindset of different actors. Interestingly, in feedback loops colleagues from the technical side, such as informatics, reported that on the one hand the change of perspective was extremely helpful for them in order to open up their mindset for ethical issues in technical development processes. However, they also reported that, on the other hand, in everyday life, there is little space and opportunity to reflect on these topics. Finally, we participated in meetings of project teams working on specific AI use cases, i.e. applications, reinforcing discussions on ELSA as part of both the research and the implementation phases of AI development. In sum, the integration of the stakeholder workshops was extremely important in order to raise awareness of ELSA aspects in the development process.

3.2 Integrating ELSA into the development and implementation of AI

The white paper focuses on three main levels: It discusses the possibility of integrating ELSA into standards software development models. It introduces the reader to legal aspects of the implementation of AI in the workplace, discussing, for example, liability issues related to the use of AI, issues of data protection, and requirements resulting from the Works Constitution Act. Finally, it develops a process-oriented approach to the human-centered use of AI that focuses on organizational development issues, discussing not only the impact of the introduction of AI on the competencies of employees, but also on effects of AI introduction on the organization of work processes within companies, and the question to what extend the use of AI is consistent with larger business objectives and culture.

We have condensed these different stages of an innovation process into a holistic phase model (cf. Alpers et al. 2023). The specific form of these stages depends on the respective context in which AI is to developed and deployed and are mutually linked. Figure 1 illustrates an idealized sequence of an AI development process which is due to be tested and adapted as part of collaborations with individual use cases and workshops within the consortium. Consequently, the phase model is meant as a framework for thinking about AI development and implementation that can and must be adapted to the concrete context of AI use. This may include, for example, the structure of the individual levels or their significance. For example, some AI applications might imply additional qualification requirements that increase the importance of the issue of retraining, or problems with the introduction of an algorithmic system can reinforce the need for accompanying organisational development, for example in the cooperation between social partners and developers. Proposals for adaption should then be discussed and, if necessary, incorporated in as inclusive

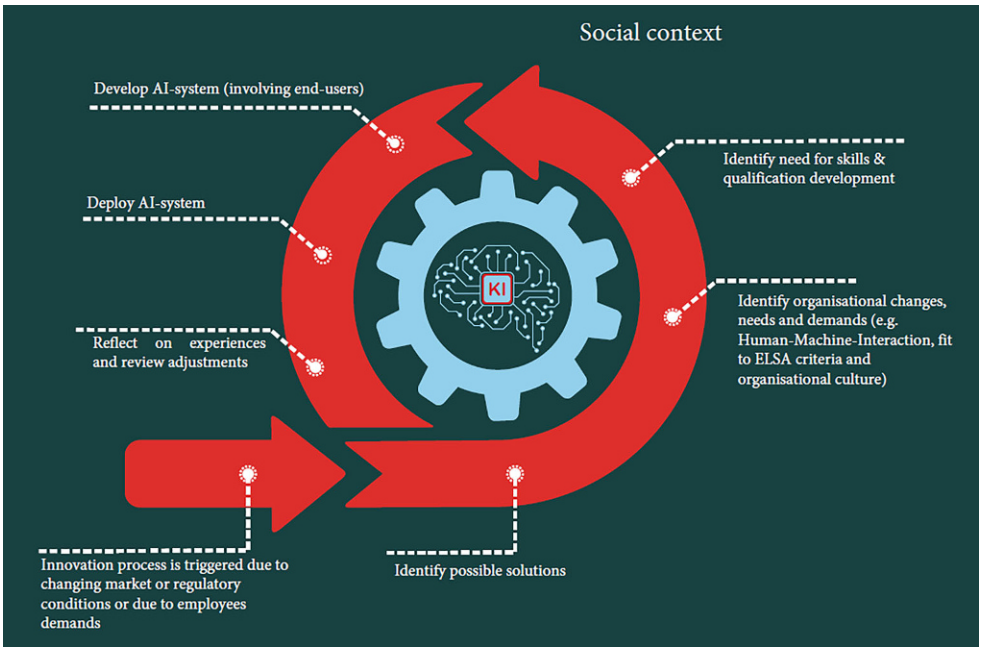

Fig. 1 Holistic model of a KI-innovation process
Abb. 1 Ganzheitliches Modell eines KI-Innovationsprozesses

a way as possible—and re-examined if necessary. In this way, the development process continues to evolve iteratively within each organization.

When it comes to ethical reflection in particular, there are certain general and highly relevant issues related to the use of AI such as risks of discrimination and performance and behavioral monitoring or the question who bears responsibility for the consequences of the use of AI, whether human agency remains intact despite the use of these systems and more. We have also found, however, that the discussion of ethical issues of AI in the workplace, especially when conducted in collaboration with practitioners, benefits greatly from being grounded as concretely as possible in the specific application context. As the following example shows, embedding ELSA-reflection into development and implementation processes can enable mutual learning processes for issues that might otherwise be dismissed as too abstract and remote.

3.2.1 Fairness in socio-technical working environments—But what type of fairness?

One of the KARL-use cases is dedicated to the development of an AI-based assistive system for personnel deployment planning in a manufacturing context. The process was designed inclusively from the beginning, both by the research partners and by the change managers of the business partner who from the get-go pursued an inclusive innovation strategy. Accordingly, the foremen responsible for personal deployment planning were involved, as well as the works council and the representative for disabled persons. Sufficient material resources were provided through the integration of the use case in KARL. The specific application area of personal planning was selected as personal planning represents a source of employee dissatisfaction in many companies, but without immediate concerns for more classical business targets such as increases in productivity.

Early on in the process, the performance criteria for the AI system were discussed. Both subjective (i.e. employee satisfaction) and normatively objective benchmarks (i.e. fairness) were identified as possibilities, with objectively "fair" personnel planning ultimately given priority. This objective then raised the question, of what fairness in personnel planning would entail. Since the working hours in the company are relatively fixed with no night shifts or weekend work, the emphasis was put on task planning, particularly in terms of the walking distance that has to be covered throughout a shift and the weight having to be lifted. An initial approach was to strive for a uniform distribution of both in distance and weight, which represented an abstract understanding of fairness as an even distribution of the burden of work. The input of the works council and the representative for disabled persons as well as inputs from us and other team members informed an adaption of the approach to take into account significant differences in individual (dis-)ability, for example based on age or disability status. The discussion therefor proceeded from weighing-up subjective and objective criteria to clarifying ethical norms of fairness in the work context, which turned out to be highly context-sensitive.

This brief example illustrates three important facets:

1. The development and introduction of new technological artifacts, such as AI systems, can serve as an opportunity

to explicate and negotiate social and ethical norms. In the preexisting mode of operation, the norms of personnel planning remained largely tacit and were applied by the foremen, leading to occasional dissatisfaction among workers. Rather than investing in social solutions such as further leadership training in conflict resolution or providing a forum for discussing the norms that guide personnel planning, and possibly risk social tensions, participation in KARL offered the opportunity to provide an AI-based solution, a technological fix (Nachtwey and Seidl 2017; Bijker and Law 1992), to this issue. Yet, the development of this technological fix itself enabled a remarkable process of negotiation of norms, that necessitated the explication of previously tacit norms that ought to guide the distribution of workload, i.e. walking distances and weights lifted, among workers. Furthermore, the organizational decision that the AI system should only assist the foremen in their planning efforts, rather than substitute them entirely, ensured that responsibility for personal planning remains with human decision-makers.

2. The inclusion of stakeholders such as the end-users (i.e. the foremen), ELSA-experts and representatives of the works council and of disabled persons constitutes an epistemic virtue (cf. Gerlsbeck and Herzog 2020): it allows to leverage additional, situated knowledge and to articulate interests, leading to software development that is more likely to meet the actual needs on the shop floor. This approach could be expanded even further, for instance by including not only the end-users of the software, the foremen, but also those subjected to the planning, i.e. ordinary workers.
3. The negotiation of social and ethical norms that guide AI development not only affects its performance metrics in the narrower sense, but also has implications for the very software architecture employed: as the implementation of negotiable and objective norms of fairness was prioritized, the development team opted against using a machine learning approach to mirror past decision making in favor of a rules-based expert system.

This process was enabled by the openness to inclusive processes of the company's coordinating change managers and by a company culture and social constitution well-attuned to technological innovation. Crucially, the resources provided by KARL (work hours, development capacities and supporting infrastructure such as access to ELSA-experts) helped dispense the oftentimes decisive orientation towards productivity gains and short-term profitability. As such, this case also illustrates the importance of institutional settings for successful human-centered technology design. The positive experiences with an inclusive, ELSA-informed, and human-centered approach to the development and implementation of AI should not distract from the fact that innovation in most work contexts is driven by the pursuit of productivity and efficiency gains, often with little regard for job satisfaction or even the well-being of workers. The primacy of short-term profitability and the dominant role of management in initiating and directing innovation can quickly lead to a neglect of ethical and social issues. It is encouraging, however, that calls for human-centered design practices and innovation processes have become more and more common in research.

We would argue that generalizing practices of human-centered technology design would require steps towards an institutionalization of processes of inclusive technology design, for example through a reform of the Works Constitution Act to guarantee more robust rights to works councils. These should not be limited to the already extensive rights to control technological innovation in the workplace and to protect workers' interests in areas such as data protection or ergonomics. Rather than being limited to such a defensive approach, which structurally casts works council members in the role of inhibitors, rather than facilitators, of innovation, these rights should include initiative rights to propose new investments and innovations. The knowledge of works councils should be harnessed and seen as a strategic resource for better innovation in the workplace.

This applies particularly when it comes to the human-centered design of AI systems, as their successful implementation often relies on a realistic picture of the work processes and working environments, with instruments such as moderated specification dialogues (cf. Schröter 2019) leading to better and more socially rounded innovation informed by the expertise of workers and their representatives. A more inclusive approach to the introduction of AI in the workplace would also help shift the focus away from generating acceptance amongst workers for technological innovation to which they are exposed with little say towards processes that help workers shape technology in their working environments. Thus, generating genuine acceptability as a capacity of the way the development and implementation process is designed to meet their needs and to appreciate their expertise seems more successful rather than mere acceptance (Fischer and Ravizza 1998; Jonas 1985).

4 Conclusion: Human-centered approaches to AI and work beyond ELSA

Coming from the human-centered approach, the ELSA-research group in the KARL consortium, has developed an internal process in order to raise awareness of the multifaced issues of humans at the workplace level and to actively integrate them into the development and implementation process of AI. As a result, we have developed

a methodological framework (guideline) that operates on two levels: on the level of the technical systems to be designed *and* on the level of a holistic development process that includes transformation issues in a long-term perspective. Thus, the guideline provides, on the one hand, social and ethical orientation for shaping future working environments. On the other hand, the implementation of an agile cooperative change management (Alpers et al. 2023) is proposed in order to reflect on quantitative and qualitative changes of future workplaces.

As the description of the research process shows, the objective of human-centered design of AI systems has been a common reference point. However, the question of what constitutes a human-centered approach to AI remained and remains challenging. In our project team, we have tried to develop a holistic approach to human-centered AI design by integrating the reflection of social, legal and ethical aspects into a common conceptual framework (Alpers et al. 2023) and operationalizing this concept in a process-oriented approach covering both new processes in software development as well as new instruments to involve workers and their representatives in the development and implementation of AI systems in the workplace.

Following this conceptual, transdisciplinary work, our stakeholder workshops and our participant observation, we will further test out the instruments we have developed, having identified challenges such as the development of common understanding between technical experts and workers to facilitate a fruitful dialogue or the question of whether software engineers should be responsiblized with engaging in ELSA-reflection as part of their own development work or whether a specialization makes sense and could be institutionalized outside of the setting of a competence center with the freedoms it affords.

At the same time, we are convinced that ethical reflection on the use of AI in work environments needs to move beyond ELSA-reflection of given AI applications in order to contribute to a truly human-centered approach to AI. After all: in a competence center focused on exploring practical applications of AI in working environments, the answer to every organizational issue is necessarily the implementation of AI systems. This is not to criticize. In contrast, the institution of such competence centers as such—especially with a technology that is developing as dynamically as AI, it can make perfect sense in terms of innovation policy to encourage research into new applications, even and particularly when some of the uses this technology might have later on is not so clear at the present stage. This institutional setting necessarily leads to a focus on technological innovation, with ELSA-reflection being a valuable but ultimately subordinate form of research.

This form of technology-push is well justified in the context of the competence center, but much less so when it comes to the introduction of AI in the economy and work environments more generally. Rather, we argue that an open-ended, unbiased and inclusive reflection of the *ends*, the organizational needs and objectives, must form the basis of any consideration of what *means* should be implemented. In particular, this reflection should take the energy intensity (cf. Rohde et al. 2021) and the economic costs associated with the introduction of AI systems in particular into account and consider alternative social innovations to avoid costly and resource-intensive over-engineering. In this, we echo the pioneering AI researcher Joseph Weizenbaum (1977) who argued already in the 1970s that not every technological potential should be exploited. Rather, a truly sovereign use of technology would imply the freedom to forgo technological solutions where social solutions might be better suited (Hengstschläger and Rat für Forschung und Technologieentwicklung 2020).

Translating this overarching insight into the KARL context leads to the conclusion that it seems highly relevant to well-define the scope of problems and dimensions that AI-systems should cope with in work environments. In doing so, the process of technological development and the process of implementation into work processes should be redesigned as an environment for negotiation between different stakeholder groups, to re-shape technological development and its social and organizational consequences, and to distribute and create reciprocal interactions between different professional groups. In summary, our framework seeks to enable a human-centered approach to AI-based work by principally empowering AI users in the workplace through robust integration. This approach should be implemented early on in the development process and the role of works councils within innovation processes should be strengthened.

As described above, the impact of AI-systems in work environments is manifold and complex with regard to organizational, social and ethical issues. In a first step, AI-system were considered primarily as a technology for further automation of production and work processes (Frey and Osborne 2017). Coming from this more traditional background, social and ethical questions deal with both the impact on employment and the long-term changes of qualification and skills. Does the implementation of AI-systems threaten existing jobs? Or in terms of inter-generational justice: does the implementation threaten future jobs? How may instruments of re-qualification and training ensure employment security? How may these instruments be developed further by enterprises and by public institutions? The topicality of these questions lies in the fact that social integration and income of most people rely fundamentally on wage labor. Technological economization increasingly raises the question of "cui bono?", who benefits from increased productivity? How can the distribution of economic

wealth be organized to enhance social and political justice? How can the welfare and income for the working population be maintained? In the light of the increasing social polarization in almost all societies, these questions and their answers are closely linked to the concept of technological and social progress (Honneth 2023; Grunwald and Hillerbrand 2021).

Another important issue concerns the development and implementation of these systems themselves, which also has a significant long scientific tradition. In addition to (new) legal problems that are intensely discussed, experiences at the workplace level shows that the use of AI-systems leads to significant changes in work organization, changes in the socio-technical setting of HMI, as well as to entirely new constellations of HMI. For instance, forms of algorithmic management can change the roles and the function of human work profoundly (Schaupp 2021). In many cases, workers are reduced to a subordinate role in this socio-technical relation or constant monitoring of the workers is implicitly established, which raises the fundamental ethical question, how the autonomy and the dignity of workers may be maintained and recognized in the future (cf. Negt 2008). This also applies to the design and development of AI-systems, where issues of individual autonomy, the appreciation of professional experience and the tacit knowledge of employees may enrich the creation and implementation processes of these systems.

As discussed above, the High-Level Group on AI recommends (European Commission 2019, p. 12), that "they [AI systems] should be designed to augment, complement and empower human cognitive, social and cultural skills." Applying these considerations to the field of work, it seems that the human-centered approach to AI-systems is still in its infancy. Ethical reflection, hereby, seems of high relevance as technological development influences notably the framework of working conditions as well as of labor markets developments. Participatory and inclusive approaches seem particularly relevant when it comes to reflecting the human side of work (Honneth 2023) and societal questions based on which norms and values AI should be developed. This question, again, may be the starting point of interrogating and reflecting "the aim for the creation of meaningful work" (European Commission 2019, p. 12).

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Albrecht S (2023) ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen. TAB-Hintergrundpapier 26. Berlin

Alpers S, Krings B-J, Welf S, Becker C, Brücklmayr J, Dreher AK, Frey P, Miriam K, Rill M, Take M, Vugrincic A, Weinreuter M (2023) KARL GF5 Leitfaden – Whitepaper zu ethischen, rechtlichen und sozialen Aspekten im Kontext von Künstlicher Intelligenz für Arbeit und Lernen. https://publikationen.bibliothek.kit.edu/1000161674/151236658. Accessed 26 Mar 2024

Alpers S, Krings B-J, Becker C, Rill M, Weinreuter M (2024) Ethische, rechtliche und soziale Aspekte (ELSA) der Gestaltung von KI-Systemen: Systematisierung der Betrachtung durch Vorgehensmodelle und Leitfäden. Z Arbeitsforsch (im Erscheinen)

Bellon J, Nähr-Wagener (2020) Interdisziplinarität, ELSI und Integrierte Forschung – aus einem Vieles und aus Vielem Eines? In: Gransche B, Manzeschke A (eds) Das geteilte Ganze. Horizontale integrierte Forschung für künftige Mensch-Technik-Verhältnisse. Springer VS, Wiesbaden, pp 37–52

Bijker WE, Law J (eds) (1992) Shaping technology; building society. Studies in sociotechnical change. MIT Press, Cambridge

Böschen S, Grunwald A, Krings B-J, Rösch C (eds) (2021) Technikfolgenabschätzung: Handbuch für Wissenschaft und Praxis. Nomos, Baden-Baden

Bundesministerium für Wirtschaft und Klimaschutz (BMWK) (2024) Plattform Industry 4.0. Online. https://www.plattform-i40.de/IP/Navigation/EN/Industrie40/WhatIsIndustrie40/what-is-industrie40.html. Accessed 26 Mar 2024

Christl W (2021) Digitale Überwachung und Kontrolle am Arbeitsplatz. Von der Ausweitung betrieblicher Datenerfassung zum algorithmischen Management? https://crackedlabs.org/dl/CrackedLabs_Christl_UeberwachungKontrolleArbeitsplatz.pdf. Accessed 27 Mar 2026

Deutscher Ethikrat (2023) Mensch und Maschine – Herausforderung durch Künstliche Intelligenz. Stellungnahme. https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf. Accessed 26 Mar 2024

El-Haouzi H, Bril V, Etienne, Krings B-J, Moniz AB (2021) Social dimensions in cps & iot based automated production systems. Societies. https://doi.org/10.3390/soc11030098

European Commission (2019) Independent expert high-level group on artificial intelligence: ethics guidelines for trustworthy AI. Brussels

Fischer J, Ravizza M (1998) Responsibility and control: a theory of moral responsibility. MIT Press, Cambridge

Frey C, Osborne MA (2017) The future of employment: how susceptible are jobs to computerisation? Technol Forecast Soc Change 114:254–280

Gerlsbeck F, Herzog L (2020) The epistemic potentials of work place democracy. Rev Soc Econ 78(3):307–330

Gethmann CF (2015) Disciplinary—interdisciplinary—transdisciplinary: a conceptual analysis. In: Gethmann CF et al (ed) Interdisciplinary research and trans-disciplinary validity claims. Springer, Berlin, Heidelberg, pp 39–60

Grunwald A (1994) Prinzip Verantwortung oder Prinzip Rechtfertigung in der Technikfolgendiskussion? Ethik Sozialwiss 51:143–145

Grunwald A (2019) Technology assessment in practice and theory. Routledge, Abingdon
Grunwald A (2022) The responsibility of researchers and engineers: codes of ethics for emerging technologies. In: Laas K, Davis M, Hildt E (eds) Codes of ethics and ethical guidelines. The international library of ethics, law and technology. In, Heidelberg, New York, Cham, pp 243–258
Grunwald A (2024) Editorial for the Special Issue: AI for decision support—What are possible futures, social impacts, regulatory options, ethical conundrums and agency constellations? J Technol Assess Theory Pract 33(1):3
Grunwald A, Hillerbrand R (eds) (2021) Handbuch Technikethik. Metzler, Stuttgart
Hengstschläger M, Rat für Forschung und Technologieentwicklung (2020) Digitaler Wandel und Ethik. Ecowin, Salzburg, München
Hirsch-Kreinsen H, Karacic A (eds) (2019) Autonome Systeme und Arbeit: Potentiale und Grenzen der Anwendung autonomer Systeme. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt. transcript, Bielefeld
Hirsch-Kreinsen H, Ittermann P, Niehaus J (eds) (2018) Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen. Nomos, Baden-Baden
Honneth A (2023) Der arbeitende Souverän. Suhrkamp, Berlin
Jonas H (1985) Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation, 4th edn. Suhrkamp, Frankfurt a.M.
Kirchschlaeger PG (2021) Digital Transformation and Ethics. Ethical Considerations on the Robotization and Automation of Society and the Economy and the Use of Artificial Intelligence. Nomos, Baden-Baden
Krings B-J, Moniz AB, Frey P (2021) Technology as enabler of the automation of work? Current societal challenges for a future perspective of work. Rev Bras Sociol 9(21):206–229
Mittelstraß J (2005) Methodische Transdisziplinarität. Technol Theor Prax 14:18–23
Müller S (2019) Das Forschungs- und Aktionsprogramm „Humanisierung des Arbeitslebens“ (1974–1989). In: Kleinöder N, Müller S, Uhl K (eds) „Humanisierung der Arbeit“. Aufbrüche und Konflikte in der rationalisierten Arbeitswelt des 20. Jahrhunderts. transcript, Bielefeld, pp 59–88
Nachtwey O, Seidl T (2017) Die Ethik der Solution und der Geist des digitalen Kapitalismus. Institute for Social Research, 2017. https://www.ifs.uni-frankfurt.de/publikationsdetails/ifs-oliver-nachtwey-und-timo-seidl-die-ethik-der-solution-und-der-geist-des-digitalen-kapitalismus.html?file=files%2FContent%2FPublikationen%2FIfS+Working+Papers%2FIfS-WP-11.pdf&fileKey=aed072b907add57b3570306ff9d375e1. Accessed 26 Mar 2024
Negt O (2008) Arbeit und menschliche Würde. Steidl, Göttingen
Nielsen M, Bryndum N, Bedsted B (2017) Organising stakeholder workshops in research and innovation—between theory and practice. J Public Deliberation 13(2):9
Owen R, Macnaghten P, Stilgoe J (2012) Responsible research and innovation: from science in society to science for society, with society. Sci Public Policy 39(6):751–760
Pereira V, Hadjielias E, Christofi M, Vrontis D (2023) A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective. Hum Resour Manag Rev 33(1):2023. https://doi.org/10.1016/j.hrmr.2021.100857
Rohde F, Wagner J, Reinhard P, Petschow U, Meyer A, Voß M, Mollen A (2021) Nachhaltigkeitskriterien für künstliche Intelligenz. Entwicklung eines Kriterien- und Indikatorensets für die Nachhaltigkeitsbewertung von KI-Systemen entlang des Lebenszyklus. Schriftenreihe des IÖW, 220/21. Berlin
Rudschies C, Schneider I, Simon J (2021) Value pluralism in the AI ethics debate—different actors, different priorities. Int Rev Inf Ethics 29/2021:
Salfer P, Furmaniak K (1981) Das Programm „Forschung zur Humanisierung des Arbeitslebens“. Stand und Möglichkeiten der Evaluierung eines staatlichen Forschungsprogramms. Mitt Arbeitsmarkt Berufsforsch 1981(3):237–245
Schaupp S (2021) Technopolitik von unten: Algorithmische Arbeitssteuerung und kybernetische Proletarisierung. Matthes & Seitz, Berlin
Schröter W (2019) Der mitbestimmte Algorithmus. Arbeitsweltliche Kriterien zur sozialen Gestaltung von Algorithmen und algorithmischen Entscheidungssystemen. In: Schröter W (ed) Der mitbestimmte Algorithmus. Gestaltungskompetenz für den Wandel der Arbeit. Mössingen, Talheimer, pp 101–150
Seyfarth R, Roberge J (eds) (2017) Algorithmuskulturen. Über die rechnerische Konstruktion der Wirklichkeit. transcript, Bielefeld
Spiekermann S (2019) Digital Ethik. Ein Wertesystem für das 21. Jahrhundert. Droemer Knaur, München
Weizenbaum J (1977) Computer power and human reason: From judgement to calculation. Freeman, San Francisco